

Kontrolliert und repräsentativ: Beispiele zur Komplementarität von Labor- und Felddaten

*Armin Falk**

Universität Bonn, IZA, Bonn, DIW Berlin und CEPR, London

Thomas Dohmen

ROA, Universität Maastricht, IZA, Bonn und DIW Berlin

Uwe Sunde

Universität St. Gallen, IZA, Bonn, CESifo, DIW Berlin und CEPR,
London

1. Einleitung

Wissenschaftliche Erkenntnis basiert maßgeblich auf der empirischen Überprüfung von Theorien, Hypothesen und Vermutungen. Hierzu stehen verschiedene Arten von Daten zur Verfügung, Labor- und Felddaten, sowie experimentelle und nichtexperimentelle Daten. Jede Datenform erlaubt spezifische Rückschlüsse auf das jeweils interessierende Phänomen und hat ihre Stärken und Schwächen.

Der wissenschaftliche Wert von Laborexperimenten für die Sozial- und Wirtschaftswissenschaften ist immens. Der Grund hierfür ist ebenso banal wie unabweisbar: Keine andere empirische Methode erlaubt ein vergleichbares Maß an Kontrolle. Im Laborexperiment kontrolliert der Wissenschaftler die (materielle) Zielfunktion der Teilnehmer, die Informationsbedingungen, alle strategisch relevanten Variablen wie etwa Marktbedingungen, ob Interaktionen wiederholt oder einmalig sind, und er stellt über die Bezahlung der Teilnehmer sicher, dass Entscheidungen reale monetäre Konsequenzen haben. Aufgrund dieser Kontrollmöglichkeiten können in der Regel exakte theoretische Prognosen hergeleitet werden. Dies wiederum ermöglicht das Testen von Theorien. Die exogene Variation von verschiedenen experimentellen

*Korrespondenzadresse: Adenauerallee 24–42, 53113 Bonn, Tel: 0228 73-9240, Fax: +49-228 73-9239, Email: armin.falk@uni-bonn.de. – Die Autoren danken Anke Becker, Simon Gächter und Benedikt Vogt für hilfreiche Kommentare.

Kontrolliert und repräsentativ: Beispiele zur Komplementarität von Labor- und Felddaten

Bedingungen ermöglicht kausale Schlussfolgerungen. Durch *random assignment* in verschiedene Zustände werden Verzerrungen durch Selektion vermieden. So lässt sich zum Beispiel die Wirkung unterschiedlicher Institutionen (Marktbedingungen, Mindestlohn etc.) kausal identifizieren, übrigens auch von Institutionen, die es außerhalb des Labors (noch gar) nicht gibt. Experimente haben die Ökonomik nachhaltig verändert. Ohne die kontrollierte Laborevidenz würden wir über zweifellos bedeutsame Phänomene wie beschränkte Rationalität, Vertrauen, Altruismus, Reziprozität, Verlustaversion, self-serving biases, Framing-Effekte, Selbstüberschätzung, sozialen Vergleich, referenzpunktabhängige Präferenzen etc. nicht oder nur sehr unvollkommen, also ohne solides empirisches Fundament sprechen. Aber sind Experimente nicht einfach unrealistisch?

Der häufig gemachte Einwand, Laborexperimente seien unrealistisch beruht auf einem doppelten Missverständnis. Erstens *sind* Entscheidungen im Experiment real. Sie werden von realen Individuen mit realen Emotionen und Überlegungen getroffen und führen zu realen monetären Konsequenzen. Zweitens ist es häufig gar nicht das Ziel eines Experiments, die Wirklichkeit genau abzubilden. Im Gegenteil: Die meisten ökonomischen Modelle sind insofern „unrealistisch“, als dass sie viele Aspekte der Realität unberücksichtigt lassen. Diese Vereinfachung ist aber von Vorteil, da sie hilft die Interaktion der wirklich relevanten Variablen zu verstehen. Ob Realitätsnähe ein wichtiges Kriterium zur Beurteilung von Experimenten ist, hängt vom Zweck des jeweiligen Experiments ab. Häufig ist es das Ziel von Laborexperimenten, Theorien zu testen. Die so gewonnene Evidenz ist relevant zur Verbesserung theoretischer Modelle und nicht für das unmittelbare Verständnis von Realität. Die einfachsten Experimente sind häufig gerade die interessantesten, denn sie liefern Einsichten in fundamentale Verhaltensweisen und allgemeine Prinzipien menschlicher Motivation.

Eines liefern Laborexperimente mit ihren typischerweise kleinen Fallzahlen und homogenen Teilnehmergruppen allerdings nicht: repräsentative Evidenz. Nun ist auch dies nicht immer ein Problem: Falsifizieren von Theorien basiert häufig gerade nicht auf spezifischen Annahmen über ihren sozioökonomischen Hintergrund. Da ökonomische Theorien typischerweise nicht nur für bestimmte Personengruppen Geltung beanspruchen, kann man sie auch mit Studenten testen. Aber wenn man an der Beschreibung von Parametern in der Bevölkerung (z.B. Diskontraten oder Risikoeinstellungen) oder am Einfluss sozioökonomischer Faktoren auf bestimmte Präferenzen oder Verhaltensweisen interessiert ist, kommt man um eine repräsentative Stichprobe nicht umhin. Eine weitere Problemstellung ergibt sich aus der Frage, inwieweit Ergebnisse aus dem Labor verallgemeinerbar sind. Dies bezieht sich vor allem darauf, ob die beobachteten Zusammenhänge auch außerhalb des Labors gelten, beispielsweise auf dem Arbeitsmarkt, und sich dort auch durchsetzen.

Es erstaunt daher, dass verschiedene Datenformen vergleichsweise selten gemeinsam zur Beantwortung einer wissenschaftlichen oder wirtschaftspolitisch relevanten Frage herangezogen werden. Das ist insofern zu beklagen,

als die Verknüpfung von unterschiedlichen Datenformen das Ausschöpfen von Komplementaritäten ermöglicht: Das Labor ermöglicht durch Kontrolle, exogene Variation von Bedingungen und zufällige Versuchsanordnungen von Untersuchungs- und Kontrollgruppen kausale Rückschlüsse, wie sie mit Felddaten in der Regel nicht möglich sind. Feld- und Surveydaten wiederum enthalten Informationen über das Verhalten in Alltagssituationen und können durch große Fallzahlen, repräsentatives Stichprobendesign und wiederholte Befragung über viele Jahre Labordaten hervorragend ergänzen.

Ziel dieses Artikels ist es, anhand von konkreten Beispielen, für die Verknüpfung verschiedener Datenformen zu werben, sowie für die damit verbundene Kooperation von Forschern mit unterschiedlichem methodischem Hintergrund. Im Folgenden werden deshalb verschiedene Einsatzmöglichkeiten von Experimenten und ihre Verbindung mit anderen Methoden diskutiert.¹ Das erste Beispiel beschreibt die Möglichkeit, Experimente auf repräsentativer Basis durchzuführen. Auf diese Weise kann zwei wichtigen Anforderungen an Daten Rechnung getragen werden: Kontrolle und Repräsentativität. Der zweite Ansatz beruht auf der Idee, Befragungsinstrumente experimentell zu validieren, und sie im Rahmen einer (existierenden) Befragung einzusetzen. Hierdurch kann die Verhaltensrelevanz von Survey-Instrumenten überprüft werden, die insbesondere unter Ökonomen häufig, ob zu Recht oder zu Unrecht, bezweifelt wird. Beim dritten Beispiel geht es darum, im Labor Zusammenhänge kausal zu identifizieren, die anschließend mit Felddaten überprüft werden können. Diese Verknüpfung von Labor- und Felddaten wird anhand von zwei Beispielen illustriert: Selbstselektion in Anreizsysteme, sowie der Einfluss unfairer Löhne auf Stress.

2. Präferenzen und IQ: Experimente mit repräsentativem Probandenkreis

Die meisten (Labor-)Experimente werden mit Studenten und Studentinnen durchgeführt. Dies erklärt sich vor allem durch die Tatsache, dass Studenten als Teilnehmer an Experimenten leicht rekrutiert werden können, und dass sie die experimentellen Regeln schnell verstehen. Die Fokussierung auf diesen Probandenkreis ist allerdings immer wieder Anlass zur Kritik. So wird beispielsweise die Verallgemeinbarkeit oder „externe Validität“ von Ergebnissen in Frage gestellt, die auf dem Verhalten eines recht beschränkten Ausschnitts der Bevölkerung beruht. Ob diese Kritik gerechtfertigt ist, hängt maßgeblich von der Fragestellung des Experiments ab: Sehr häufig geht es im Experiment um das Überprüfen (im strikten Sinne: Falsifizieren) von ökonomischen Theorien. Diese Theorien beanspruchen in aller Regel Geltung zur Beschreibung

1. Dieser Beitrag beruht auf einem Vortrag bei der Tagung des Vereins für Socialpolitik 2008 in Graz. Die hier diskutierten Ideen und Resultate beruhen zum größten Teil auf Arbeiten von Thomas Dohmen, Armin Falk, David Huffman, Jürgen Schupp, Uwe Sunde und Gert Wagner.

Kontrolliert und repräsentativ: Beispiele zur Komplementarität von Labor- und Felddaten

menschlichen Verhaltens, können also im Prinzip mit allen Menschen untersucht werden. Uns sind keine Theorien geläufig, die nur für Nicht-Studenten Geltung beanspruchen, also auch nur mit Nicht-Studenten überprüft werden könnten. Zudem sind Ökonomen häufig an qualitativen Effekten interessiert, die über Treatment-Unterschiede ermittelt werden können und weniger an quantitativen Effekten.² Es gibt aber Fragestellungen, die zweifellos einen weniger restriktiven Probandenkreis erfordern. Ein besonders augenscheinliches Beispiel ist die Frage nach der Verteilung von Präferenzparametern in der Bevölkerung. Die Ermittlung der Verteilungen von Risikobereitschaft, Zeitpräferenzen, oder Freizeitpräferenzen, oder der Einstellungen gegenüber anderen (sozialen Präferenzen) oder anderer Persönlichkeitsmerkmale erfordern Beobachtungen auf repräsentativer Ebene.

Ein Beispiel für die Verbindung von Experimenten und repräsentativen Stichproben ist eine Studie von Dohmen et al. (2007). Darin wird der Zusammenhang zwischen kognitiven und nicht-kognitiven Fähigkeiten untersucht, und zwar konkret im Hinblick auf die Fragestellung, ob es einen signifikanten Zusammenhang zwischen Intelligenz, gemessen mit Hilfe von Testverfahren für verschiedene Bereiche kognitiver Fähigkeiten, und ökonomisch relevanten Präferenzen, wie Zeit- und Risikopräferenzen, gibt. Die Erforschung dieses Zusammenhangs erfordert repräsentative Daten: Verließe man sich alleine auf Studenten, würde die Untersuchung mögliche Variationen dieses Zusammenhangs in verschiedenen Bereichen der Intelligenzverteilung vernachlässigen. Zudem könnte über den Zusammenhang zwischen Intelligenz und Präferenzen in anderen sozio-demographischen Gruppen nur spekuliert werden.

In den Wirtschaftswissenschaften wird in der Regel angenommen, dass kognitive Fähigkeiten und Präferenzen unabhängig voneinander sind. Hinweise aus der kognitiven Psychologie sowie aus der neuroökonomischen Forschung legen allerdings die Vermutung nahe, dass höhere kognitive Fähigkeiten positiv korreliert sind mit Geduld und der Bereitschaft, Risiken einzugehen (Dohmen et al., 2007). Die Klärung dieser Frage ist über die Grundlagenforschung hinaus auch für angewandte Fragen von Bedeutung. Zum Beispiel würde das Vorliegen einer systematischen Korrelation von Intelligenz und Risikobereitschaft bzw. Geduld bedeuten, dass der direkte Einfluss von Intelligenz auf ökonomische Ergebnisvariablen, wie etwa Einkommen, überschätzt wird, wenn nicht gleichzeitig für Präferenzen kontrolliert wird. Letztere tragen schließlich unabhängig zum ökonomischen Erfolg bei.

An der Studie nahmen etwa 1000 Probanden teil. Die Auswahl der Probanden wurde von Infratest mit Hilfe der so genannten Random Route Methode vorgenommen (siehe Dohmen et al., 2005, für Details). Dieses Sampling-Verfahren wird häufig auch bei der Durchführung von großen Befragungsstudien verwendet. Die Befragungen und Experimente fanden nicht im Labor,

2. Für eine Diskussion zur Methode der Feld- und Laborexperimente siehe Falk und Heckman (2009).

sondern im „Feld“ statt: sie wurden bei den Probanden zu Hause und computergestützt durchgeführt (computer assisted personal interview).

Die Intelligenz der Probanden wurde mit Hilfe zweier Untermodule des sogenannten Hamburg-Wechsler Intelligenztest für Erwachsene gemessen, dem wichtigsten und am häufigsten verwendeten Instrument zur Bestimmung von Intelligenzquotienten (Borghans et al., 2008). Die beiden Untermodule waren ein Symbol-Zuordnungstest-Test, bei dem in 90 Sekunden Symbole und Zahlen zugeordnet werden, sowie ein Wortschatztest, bei dem in 90 Sekunden so viele Tiernamen wie möglich genannt werden müssen (Lang, 2005).

Risiko- und Zeitpräferenzen wurden mit Hilfe experimenteller Methoden erhoben.³ Die Ermittlung der individuellen Risikopräferenzen erfolgte dabei mit Hilfe einer Tabelle. Teilnehmer am Experiment mussten sich in jeder Zeile zwischen jeweils zwei Optionen entscheiden, einer Lotterie (300 Euro oder 0 Euro, jeweils mit 50 Prozent Wahrscheinlichkeit) oder einer sicheren Auszahlung, die in jeder Zeile um 10 Euro bis zu einem Betrag von 190 Euro ansteigt. Da der Erwartungswert der Lotterie 150 Euro beträgt, wählen die meisten Teilnehmer in Zeilen mit geringen sicheren Auszahlungen die Lotterie. Je höher aber die sichere Auszahlung, desto mehr Teilnehmer entscheiden sich für die sichere Auszahlung. Die Zeile, in der ein Teilnehmer von der Lotterie zur sicheren Auszahlung wechselt, informiert uns über seine Risikopräferenz, bzw. das Sicherheitsäquivalent: An dieser Stelle ist er (nahezu) indifferent zwischen der Lotterie und einer sicheren Zahlung. Nachdem ein Teilnehmer für jede Zeile eine Entscheidung getroffen hat, wird per Zufall bestimmt welche Zeile tatsächlich zur Auszahlung kommt. Hierdurch wird sichergestellt, dass jede Entscheidung potenziell auszahlungsrelevant ist und der Experimentteilnehmer daher einen Anreiz hat, seine „wahre“ Präferenz zu offenbaren. Die Erhebung von Zeitpräferenzen erfolgte analog. Hier bestand jeweils die Wahl zwischen einer Auszahlung von 100 Euro „heute“ und einer ansteigenden Auszahlung in „einem Jahr“. Alle Geldbeträge wurden per Scheck zugesandt. Neben der Erhebung von Präferenzen und IQ wurde ein ausführlicher Fragebogen zum sozioökonomischen Hintergrund (angelehnt am Sozioökonomischen Panel (SOEP)) sowie psychologischen Hintergrundvariablen beantwortet.

Das Hauptresultat ist in Abbildung 1 dargestellt.⁴ Sie zeigt, dass Personen mit höherer gemessener Intelligenz sowohl geduldiger als auch risikotoleranter sind. Personen mit relativ höheren kognitiven Fähigkeiten springen erst bei höheren sicheren Auszahlungen von der Lotterie auf die sichere Auszahlung, d.h. ihr Sicherheitsäquivalent ist höher. Umgekehrt ist für diese Personen eine relativ niedrige Rendite ausreichend, um einige Zeit auf die Auszahlung zu warten. Dieser Befund zeigt, dass kognitive und nicht-kognitive Fähigkeiten in einer systematischen Wechselbeziehung stehen. Das in Abbildung 1

3. Vgl. zum Beispiel Andersen et al. (2008).

4. Wir präsentieren in dieser Arbeit Resultate vorwiegend mittels Grafiken. Statistische Tests, Regressionen etc. finden sich in den zitierten Originalarbeiten. Alle berichteten Ergebnisse sind statistisch signifikant.

Kontrolliert und repräsentativ: Beispiele zur Komplementarität von Labor- und Felddaten

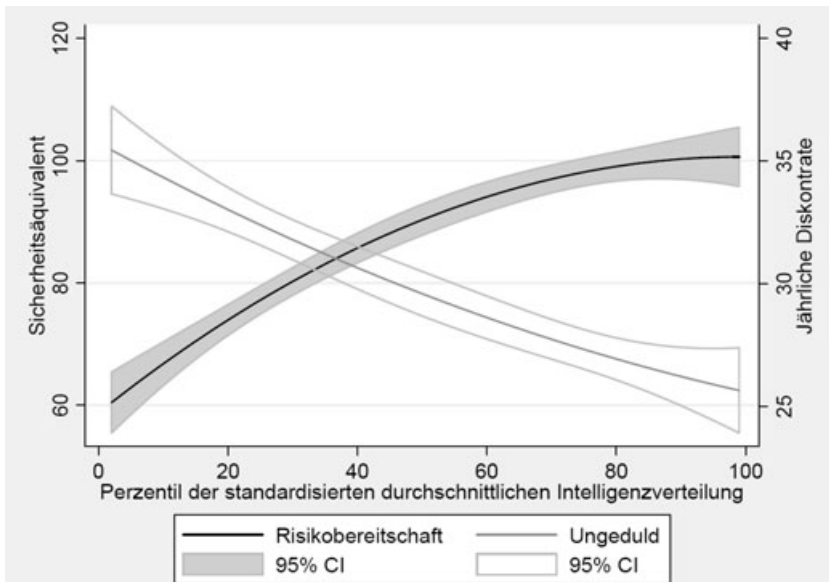


Abbildung 1 Zusammenhang zwischen Kognitiver Fähigkeit, Risikoeinstellung und Geduld

Quelle: Dohmen et al. (2005)

gezeigte Resultat ist robust und bleibt statistisch signifikant, wenn man für verschiedene Hintergrundvariablen, wie beispielsweise Alter, Geschlecht und Einkommen kontrolliert. Es zeigt sich weiterhin, dass der Zusammenhang von IQ und Präferenzen gruppenspezifisch ist. So ist der Zusammenhang beispielsweise für Männer stärker als für Frauen.

3. Experimentelle Validierung von Befragungsinstrumenten

Das im vorigen Abschnitt diskutierte Beispiel veranschaulicht, dass die Durchführung von Experimenten und die Verwendung von repräsentativen Samples sich keineswegs ausschließen. Allerdings ist diese Verfahrensweise relativ teuer und trotz der über 1000 Teilnehmer wären höhere Fallzahlen häufig wünschenswert, um detaillierte Analysen vornehmen zu können. Das gilt, wenn man z.B. neben Verteilungen auch an den Determinanten von Präferenzen interessiert ist, und an ihrem Erklärungsgehalt für ökonomische Handlungen und Ergebnisse (Löhne, Berufswahl, Partnerwahl etc.). Aus diesem Grund haben Dohmen et al. (2005) für die Risikopräferenzmessung ein Verfahren verwendet, bei dem ein Fragebogeninstrument zur Messung von Risikoeinstellungen experimentell validiert wurde, welches im Rahmen des Sozioökonomi-

schen Panels (SOEP) verwendet wird.⁵ Im SOEP werden jährlich etwa 22.000 Personen in Deutschland auf repräsentativer Basis zu Hause befragt (Schupp und Wagner 2002, Frick, Schupp und Wagner, 2007, Wagner et al., 2008). Das SOEP ist eines der größten und ältesten Panels und zeichnet sich neben der großen Fallzahl auch dadurch aus, dass nicht nur Einzelpersonen sondern Haushalte befragt werden. Zudem verfügt es neben ausführlichen Informationen zu sozioökonomischen Merkmalen der Befragten auch über Informationen zu Präferenzen, Persönlichkeit und Einstellungen. Die Komplementarität von Experiment und Survey besteht in diesem Fall also darin, dass sichergestellt wird, dass das verwendete Survey-Instrument verhaltensrelevant ist (es prognostiziert Entscheidungen im bezahlten und kontrollierten Experiment) und gleichzeitig auf sehr breiter Basis in einem bestehenden Panel administriert werden kann.

Die Risikofrage, die wir experimentell validiert haben, wurde im Sozioökonomischen Panel (SOEP) in der 2004er Welle verwendet. Sie lautet:

„Wie schätzen Sie sich persönlich ein: Sind Sie im Allgemeinen ein risikobereiter Mensch oder versuchen Sie, Risiken zu vermeiden?“

Bitte kreuzen Sie ein Kästchen auf der Skala an, wobei der Wert 0 bedeutet: „gar nicht risikobereit“ und der Wert 10: „sehr risikobereit“.

Mit den Werten dazwischen können Sie Ihre Einschätzung abstufen.“

Diese eher allgemein gehaltene Frage wird in der 2004-er Welle um kontextspezifische Risikofragen ergänzt. Diese erfassen konkret die Risikobereitschaft in den Lebensbereichen Gesundheit, Verkehr, finanzielle Angelegenheiten, Sport und Freizeit und Beruf.

Der Nachteil von Befragungsinstrumenten, vor allem in den Augen vieler Ökonomen, besteht darin, dass sie kein Verhalten messen. Die Antworten sind nicht mit wirtschaftlichen Anreizen versehen und es ist daher nicht auszuschließen, dass die Teilnehmer beliebige Antworten geben, etwa, weil sie strategisch antworten (soziale Erwünschbarkeit etc.) oder Fragen einfach nicht ernst nehmen. Zudem stellen Fragebogenszenarien erhebliche Anforderungen an das Abstraktionsvermögen der Teilnehmer, und Antworten hängen möglicherweise vom Kontext ab, in dem die Fragen gestellt wurden. Die latente Geringschätzung von Fragebogenitems führt allerdings dazu, dass sich Ökonomen selbst einer sehr reichhaltigen Datenquelle berauben. Die experimentelle Validierung könnte helfen, die Skepsis zu verringern, wenn gezeigt werden kann, dass Antworten auf Fragen auch tatsächlich Verhalten prognostiziert.

Um zu ermitteln, ob Antworten auf die oben dargestellte Risikofrage Risikoverhalten unter kontrollierten Bedingungen prognostiziert, führten Dohmen et al. (2005) das oben beschriebene Experiment in einer Stichprobe bestehend aus 450 Teilnehmern durch, die repräsentativ für die Erwachsenenbevölkerung in Deutschland ist. Die Stichprobe von Teilnehmern am Experiment

5. Für ein ähnliches Verfahren im Kontext von Vertrauen, siehe Fehr et al. (2003).

Kontrolliert und repräsentativ: Beispiele zur Komplementarität von Labor- und Felddaten

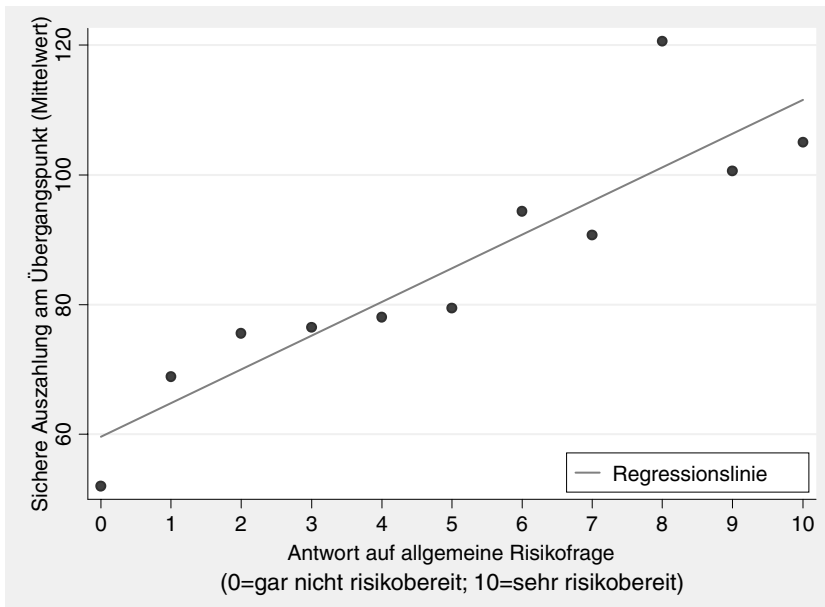


Abbildung 2 Zusammenhang zwischen Risikoverhalten (vertikale Achse: Sicherheitsäquivalent im Experiment) und Antwort auf allgemeine Risikofrage
Quelle: Dohmen et al. (2005)

unterscheidet sich in wesentlichen sozioökonomischen Merkmalen kaum von den Befragten des SOEP. Die Teilnehmer am Risiko-Experiment hatten in einem vorausgehenden Fragebogen auch die allgemeine Risikofrage beantwortet. Wie Abbildung 2 zeigt, sind die Antworten auf die Risikofrage und Entscheidungen im Experiment korreliert. Teilnehmer, die angeben, dass sie eher bereit sind Risiken einzugehen, wechseln später (bei höheren Werten) zur sicheren Auszahlung als Teilnehmer, die wenig risikobereit sind. Mit anderen Worten können Antworten auf die Risikofrage verwendet werden, um tatsächliches kontrolliertes Risikoverhalten zu prognostizieren. Allgemein lässt sich festhalten, dass durch die experimentelle Validierung die Verlässlichkeit eines Survey-Instruments deutlich erhöht werden kann, wenn sich zeigt, dass es Verhalten unter experimentellen Bedingungen erklärt.

Mit Hilfe des experimentell validierten Maßes können nun zahlreiche Fragestellungen im Rahmen des SOEP angegangen werden. In verschiedenen Arbeiten untersuchen wir z.B. die folgenden Fragen: Wie risikobereit ist die deutsche Bevölkerung? Welches sind wichtige Determinanten von Risikoeinstellungen? Woher kommen unsere Präferenzen? Werden Risikoeinstellungen vererbt? Gibt es *Assortative Mating* im Hinblick auf Risikoeinstellungen? Sind Risikoeinstellungen kontextspezifisch? Können Risikoeinstellungen wichtige ökonomische Entscheidungen erklären, wie z.B. Migration, Berufswahl oder

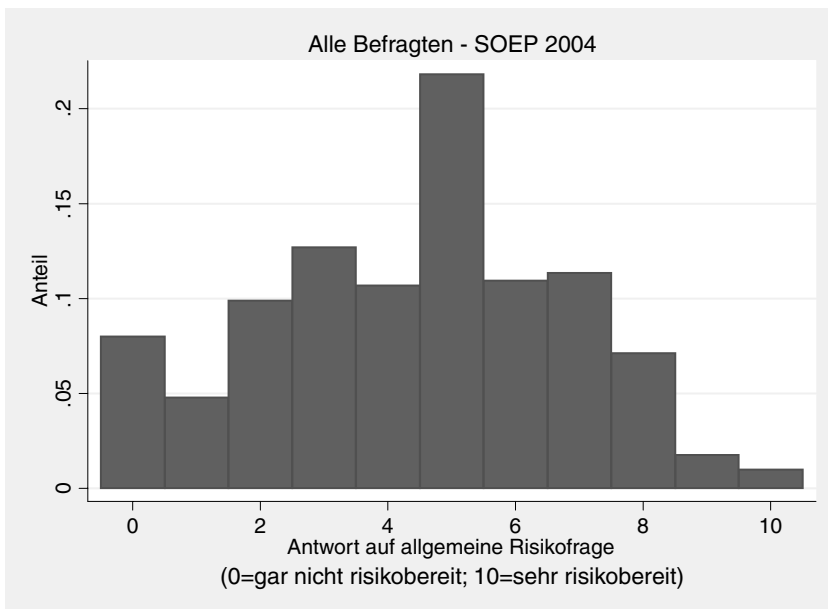


Abbildung 3 Verteilung der Antwort auf eine allgemeine Risikofrage
Quelle: Dohmen et al. (2005)

Aktienbesitz? Antworten auf diese Fragen erfordern einen repräsentativen Datensatz mit Panel-Struktur, Haushaltsdaten und ausführliche Informationen über sozioökonomischen Hintergrund und ökonomische Entscheidungen.

Im Folgenden beantworten wir die gestellten Fragen mit Hilfe einiger Abbildungen. Abbildung 3 zeigt, wie Antworten auf die Risikofrage in Deutschland verteilt sind. Die Abbildung belegt zum einen, dass die meisten Menschen in Deutschland eher wenig bereit sind, Risiken einzugehen. Die Mehrheit gibt auf der 11-Punkte Skala Werte von kleiner oder gleich 5 an.⁶ Zum anderen zeigt die Grafik eine stark ausgeprägte Heterogenität in der Risikobereitschaft: Menschen sind in Bezug auf ihre Risikoeinstellungen sehr unterschiedlich. Dieser Befund mag vielleicht trivial erscheinen; man sollte sich aber daran erinnern, dass in den (makro-) ökonomischen Modellen in der Regel von einem repräsentativen Agenten ausgegangen wird. Die experimentell gestützte empirische Forschung mit repräsentativen Stichproben kann dabei helfen, über Präferenzparameter zu informieren, um somit den Erklärungsgehalt von Modellen und Simulationen zu verbessern.

Angesichts der Heterogenität stellt sich die Frage, ob es exogene Faktoren gibt, die einen Teil dieser Heterogenität erklären können. Wir untersuchen u.

6. Für eine quantitative Interpretation der Antworten auf die Risikofrage, etwa im Sinne von Koeffizienten der konstanten relativen Risikoaversion (CRRA), siehe Dohmen et al. (2005).

Kontrolliert und repräsentativ: Beispiele zur Komplementarität von Labor- und Felddaten

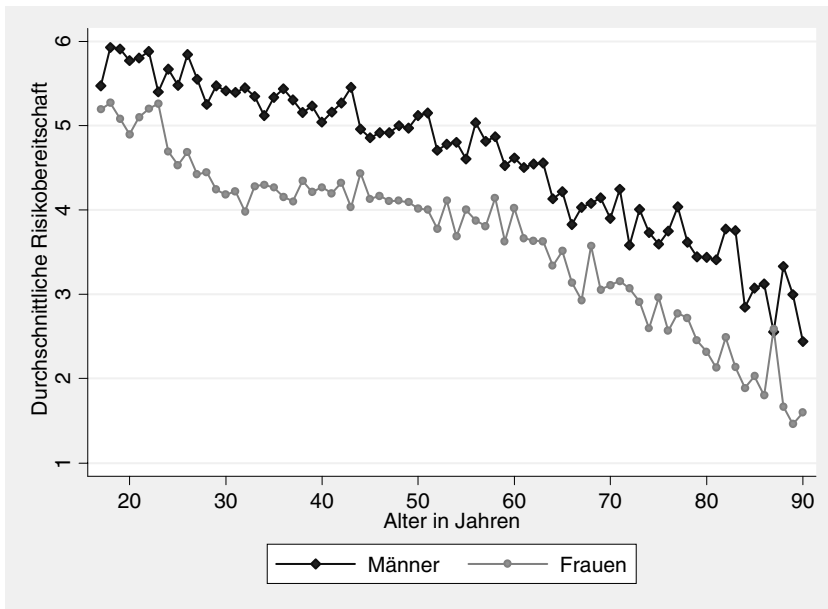


Abbildung 4 Risikoeinstellung (vertikale Achse: Antworten auf die Risikofrage) nach Alter und Geschlecht

Quelle: Dohmen et al. (2005)

a. die Rolle von Alter, Geschlecht und Körpergröße. Abbildung 4 zeigt, dass ältere Menschen signifikant weniger risikobereit sind als jüngere. Zudem zeigt die Abbildung, dass Männer bereit sind, Risiken einzugehen als es bei Frauen der Fall ist. Der Alters- und Geschlechtereffekt ist quantitativ bedeutsam und robust, wie multiple Regressionen mit Kontrollen für Einkommen, Region, Bildungshintergrund, Vermögen, Religionszugehörigkeit etc. zeigen.

Neben Alter und Geschlecht hat auch die Körpergröße einen signifikanten Einfluss auf die Risikobereitschaft. Wie Abbildung 5, getrennt für Frauen und Männer, zeigt, sind größere Menschen signifikant risikobereiter als kleinere Menschen. Auch dieser Effekt ist stark und robust. Die verschiedenen Determinanten zur Risikoeinstellung können wichtige Verhaltensunterschiede implizieren. Geschlechtsspezifische Risikovorlieben können über Selbst-Selektion eine Rolle bei Lohnunterschieden und geschlechtsspezifischer Berufswahl spielen. Da Berufe mit variabler Entlohnung im Durchschnitt besser bezahlt sind, aber Mitarbeiter auch zusätzlichen Risiken aussetzen, sollten Frauen weniger häufig in variabel entlohnenden und besser bezahlten Jobs anzutreffen sein (s.u.). Frauen finden sich beispielsweise überproportional stark im öffentlichen Dienst, der durch relativ sichere aber auch relativ geringe Einkommen gekennzeichnet ist. Die Tatsache, dass die Körpergröße einen Einfluss auf die Risikobereitschaft hat, könnte eine Teilerklärung für das empirische Faktum

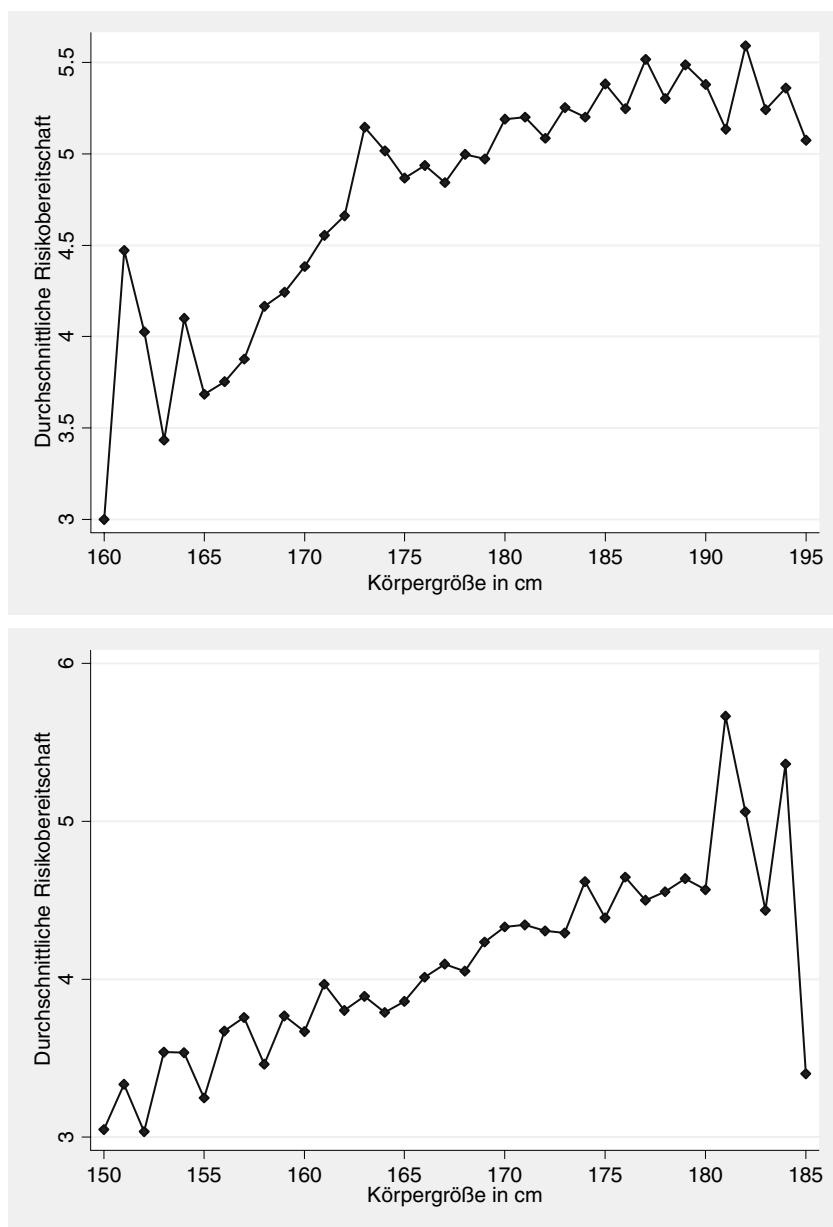


Abbildung 5 Risikoeinstellung (vertikale Achse: Antworten auf die Risikofrage) aufgeteilt nach Körpergröße und Geschlecht (oberes Panel Männer, unteres Panel Frauen)

Quelle: Dohmen et al. (2005)

Kontrolliert und repräsentativ: Beispiele zur Komplementarität von Labor- und Felddaten

liefern, dass größere Menschen erfolgreicher sind und über höhere Einkommen verfügen (Persico et al., 2004). Der systematische Alterseffekt auf die Risikobereitschaft hat weit reichende Konsequenzen für alternde Gesellschaften. Wenn die Bevölkerung älter und damit risikoscheuer wird, sollte dies die Investitionsbereitschaft, individuelle Finanzportfolios und politische Wahlentscheidungen systematisch beeinflussen.

Eine weitgehend unbeantwortete Frage ist die nach der Herkunft von Präferenzen. Eine Möglichkeit besteht darin, dass Präferenzen von Eltern an ihre Kinder weitergegeben werden. Die Tatsache, dass das SOEP über Haushaltsdaten verfügt, ermöglicht es, die intergenerationale Weitergabe von Präferenzen zu studieren. In Dohmen et al. (2006) analysieren wir diese Frage für Risiko- und Vertrauenspräferenzen im Detail. Ein wichtiges Resultat ist in Abbildung 6 abgebildet. Sie zeigt, basierend auf den Daten des SOEP, getrennt für Väter und Mütter, dass es eine sehr starke Korrelation in den Risikoeinstellungen von Eltern und Kindern gibt.

Die intergenerationale Weitergabe ist natürlich umso stärker, je ähnlicher die Präferenzen der Ehepartner sind. Sind die Partner verschieden, heben sich die entgegen gesetzten Einflüsse auf; sind sich die Eltern ähnlich, wird die Präferenzweitergabe verstärkt. Ob Partner in Bezug auf ihre Präferenzen ähnlich sind oder nicht, ist letztlich eine empirische Frage. In der theoretischen Literatur wurde einerseits zugunsten von negativem *Assortative Mating* argumentiert (Chiappori und Reny, 2006), also prognostiziert, dass aus Gründen der besseren Risikostreuung Partner mit jeweils unterschiedlichen Risikoneigungen bevorzugt werden. Andererseits gibt es theoretische wie empirische Hinweise auf positives *Assortative Mating*, also darauf, dass die Wahl oft auf Partner mit ähnlichen Attributen und Einstellungen fällt (Sichelstiel und Söllner, 2004).⁷ Basierend auf Daten des SOEP (über 7000 Paare) zeigt Abbildung 7, dass es in Bezug auf Risikopräferenzen ein stark ausgeprägtes und signifikantes positives *Assortative Mating* gibt. Die Risikovorlieben von Ehepartnern sind stark positiv korreliert, ein Zusammenhang, der robust bleibt, wenn für verschiedenste Einflussgrößen kontrolliert wird.

Neben der Frage, was unsere Risikoeinstellungen beeinflusst, woher sie kommen und wie ähnlich wir unseren Partnern sind, lässt sich zeigen, dass die validierte Risikofrage des SOEP auch zahlreiche relevante ökonomische Verhaltensweisen und Ergebnisse erklären kann. Hierzu zählen z.B. Lohnhöhe, Lebenszufriedenheit, Aktienbesitz, die Frage, ob man im öffentlichen Sektor oder der Privatwirtschaft arbeitet, selbständig ist, raucht oder Sport treibt (Dohmen et al., 2005). Die weit reichenden Erklärungspotentiale von Präferenzheterogenität zeigen auch die Arbeiten von Bonin et al. (2007) und Jaeger et al. (2009). Erstere belegt, dass Risikopräferenzen die Selbstselektion in Berufszweige beeinflussen, wobei riskantere Berufszweige mit hoher Lohnvariabilität systematisch von relativ risikobereiten Arbeitnehmern bevorzugt werden.

7. Die Alltagsweisheit hilft hier nicht weiter: "Gleich und gleich" ziehen sich ja bekanntlich genau so an, wie "Gegensätze"!

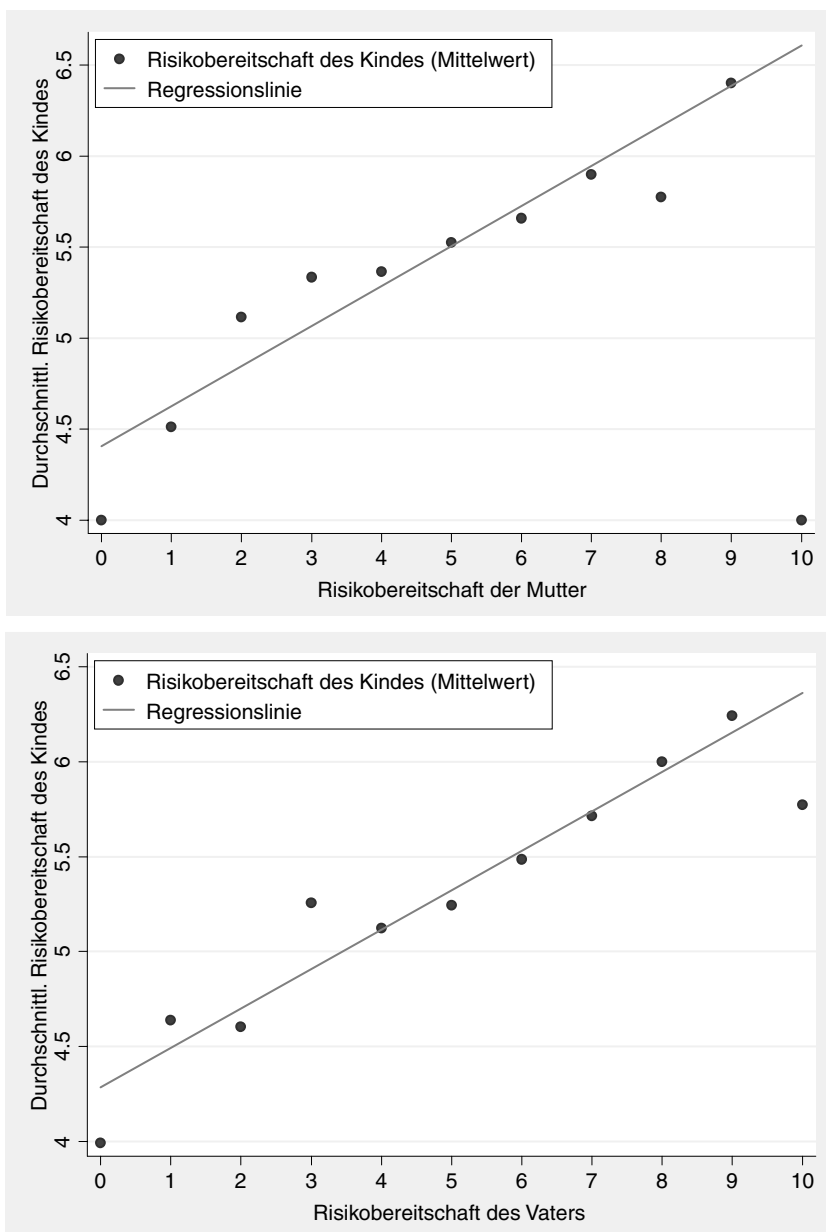


Abbildung 6 Risikoeinstellung (Antworten auf die Risikofrage) von Kind und Elternteil

Quelle: Dohmen et al. (2006)

Kontrolliert und repräsentativ: Beispiele zur Komplementarität von Labor- und Felddaten

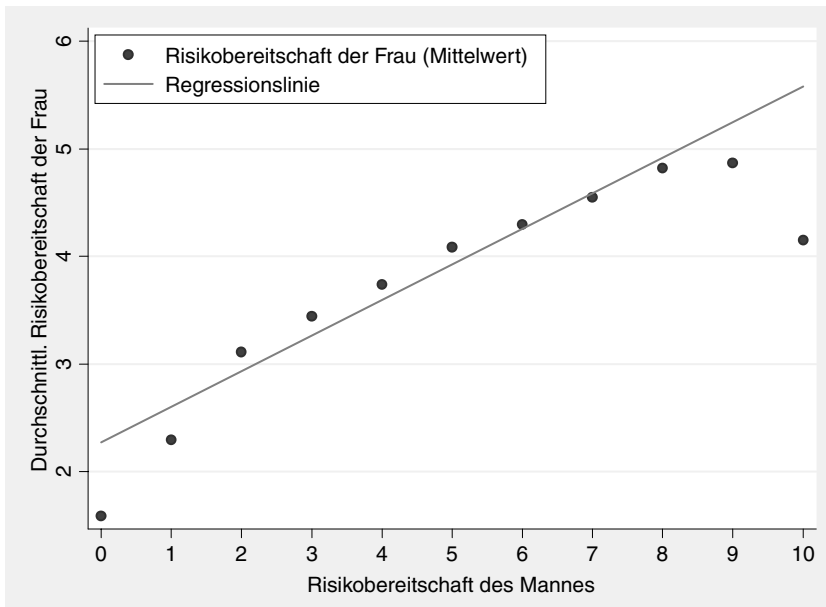


Abbildung 7 Risikoeinstellungen (Antworten auf die Risikofrage) der Ehepartner
Quelle: Dohmen et al. (2006)

Letztere zeigt, dass Migrationsentscheidungen innerhalb von Deutschland systematisch durch Risikopräferenzen bestimmt werden. Auch im internationalen Vergleich von Risikopräferenzen zeigen sich interessante Ergebnisse. So können Fehr et al. (2003) zeigen, dass US Amerikaner systematisch risikobereiter sind als Deutsche, und zwar unter Verwendung des gleichen Risikomaßes wie im SOEP. Der Unterschied ist beachtlich und entspricht etwa einer halben Standardabweichung der Risikoeinstellungen in Deutschland. Internationale Unterschiede in Risikopräferenzen könnten einen Beitrag zur Erklärung unterschiedlicher ökonomischer Entwicklungen und der Herausbildung von Institutionen liefern.

4. Laborexperimente liefern Hypothesen, die im Feld überprüft werden

Eine dritte Möglichkeit, Experimentaldaten und Felddaten gemeinsam zu nutzen, besteht darin, im Experiment Hypothesen zu gewinnen, die anschließend mit Felddaten überprüft werden. Die Dividende einer solchen Übung ist eine doppelte: Zum einen lässt sich die Verallgemeinerbarkeit von experimentell gewonnenen Ergebnissen mit Felddaten überprüfen. Zum anderen können Experimente dabei helfen, kausale Interpretationen, die mit Felddaten notorisch

schwierig sind, mit Hilfe von Labordaten, die eindeutig kausal interpretierbar sind, zu stützen. Wir werden nachfolgend zwei Beispiele diskutieren, die beide Aspekte verdeutlichen.

Im ersten Beispiel untersuchen Dohmen und Falk (2006) mit Labordaten die Selbstselektion von Arbeitnehmern in unterschiedliche Anreizsysteme bevor sie fragen, ob die im Labor beobachtbaren Selektionsmuster auch auf dem deutschen Arbeitsmarkt sichtbar sind. Anreizverträge haben grundsätzlich zwei Effekte: den eigentlichen Anzeizeffekt sowie einen potentiellen Selektionseffekt, d.h. die Tatsache, dass Anreizbedingungen in Organisationen mitentscheiden, *wer* in einer Organisation arbeitet. Mit anderen Worten entscheiden Anreize über die Charakteristika einer Belegschaft, z.B. in Hinblick auf ihre Produktivität, Präferenzen und Persönlichkeitsmerkmale. Während Anzeizeffekte in der Literatur vielfach diskutiert wurden, ist unser Wissen um Selektionseffekte recht beschränkt. Dies ist insofern überraschend, da Selektionseffekte die Zusammensetzung der Belegschaft einer Organisation bestimmen und somit für deren Erfolg entscheidend sein können.

Die Teilnehmer am Experiment mussten, um Geld zu verdienen, einfache Multiplikationsaufgaben lösen (einstellige mal zweistellige Zahlen). Dabei konnten sie wählen, ob sie nach einer fixen Entlohnung arbeiten wollten oder nach einer variablen. Wer die fixe Entlohnung wählte, erhielt einen Geldbetrag unabhängig von der Anzahl der richtig gelösten Aufgaben. Bei der variablen Entlohnung stieg die Bezahlung in der Anzahl richtig gelöster Aufgaben. Letztere konnte entweder in Form einer einfachen Stücklohnbezahlung, oder in Form eines Turnier- bzw. eines Teamanreizes gestaltet sein. Das Experiment hatte jeweils drei Phasen: zuerst wurde die individuelle Produktivität der Teilnehmer gemessen. Anschließend trafen die Teilnehmer die Entscheidung, ob sie eine fixe oder variable Entlohnung bekommen und begannen dann zu arbeiten. In der dritten Phase schließlich wurden verschiedene individuelle Charakteristika bestimmt, v.a. Präferenzen (Risikobereitschaft, Reziprozität, Vertrauen), sowie Persönlichkeitsmerkmale (Selbstüberschätzung, Big Five).

Die Resultate zeigen das folgende Bild: Die Wahrscheinlichkeit, dass ein Teilnehmer die variable Entlohnung wählt, ist höher, je risikobereiter und produktiver er ist; sie ist geringer, je stärker reziprok er ist. Zudem zeigt sich, dass Frauen tendenziell häufiger die fixe Entlohnung wählten. Die beschriebenen Selektionseffekte sind kausal eindeutig; sind sie aber auch in der Lage, Selektionsentscheidungen auf dem deutschen Arbeitsmarkt zu prognostizieren? Immerhin ist die Laborsituation hoch artifiziell, und alle Entscheidungen wurden von Studenten und Studentinnen getroffen, von Probanden also, die für den Arbeitsmarkt nicht repräsentativ sind. Dohmen und Falk (2006) untersuchen daher mit Hilfe des SOEP, ob die im Labor beobachteten Selektionsmuster auch für den deutschen Arbeitsmarkt insgesamt gelten. Das SOEP stellt die hierfür notwendigen Informationen bereit.

Kontrolliert und repräsentativ: Beispiele zur Komplementarität von Labor- und Felddaten

Arbeitnehmer wurden im SOEP darüber befragt, ob ihre Entlohnung leistungsabhängig oder fix ist.⁸ Zudem wurden Fragen zur Risikobereitschaft (s.o.), zu Reziprozität und Vertrauen (Dohmen et al., 2008, 2009) gestellt. Zu Reziprozität wurden sechs Fragen gestellt, die die Bereitschaft abfragen, faires oder kooperatives Verhalten zu belohnen und unfaires Verhalten zu bestrafen. Bei den Vertrauensfragen wurde ermittelt, inwiefern Personen glauben, dass man anderen vertrauen kann. Die Ergebnisse sind in Tabelle 1 dargestellt. Spalte 1 zeigt, dass die Vorzeichen der untersuchten Variablen exakt so sind, wie aus den Labordaten prognostiziert. Die Wahrscheinlichkeit, dass ein Arbeitnehmer sich für leistungsabhängige Bezahlung entscheidet, ist umso höher, je risikobereiter er ist und geringer für Frauen und reziproke Arbeitnehmer. Die Effekte bleiben signifikant wenn man für Produktivität (Proxy: Schuljahre) kontrolliert (Spalte 2). Wie im Experiment ist der Koeffizient für Produktivität positiv. Auch weitere Kontrollen (Spalte 3) lassen die Zusammenhänge weitgehend unverändert. Alter und Teilzeit verringern die Wahrscheinlichkeit leistungsabhängiger Bezahlung, während die Wahrscheinlichkeit positiv von der Firmengröße abhängt. Insgesamt zeigt sich also, dass Selektionsmuster, die unter kontrollierten Bedingungen im Labor auftreten, auch für Entscheidungen am deutschen Arbeitsmarkt relevant sind.

Im zweiten Beispiel gehen wir den umgekehrten Weg, von Felddaten zu Labordaten. Es existiert umfassende Literatur zum Einfluss von als „ungerecht“ wahrgenommener Entlohnung auf Stress und kardiovaskuläre Erkrankungen (Siegrist, 1996), allerdings ist die kausale Interpretation problematisch. Es ist gleichermaßen plausibel anzunehmen, dass Löhne einen Einfluss auf Stress und Gesundheit haben, oder dass gesundheitliche Beeinträchtigungen einen Einfluss auf die Entlohnung haben. Die Klärung, ob eine Kausalität von Unfairness auf Stress vorliegt, ist von großer Bedeutung, nicht zuletzt weil Stress erwiesenermaßen zu Herz-Kreislauferkrankungen führt. Experimente können an dieser Stelle zeigen, ob ein Wirkungskanal kausal existiert.

Im Experiment von Falk et al. (2008) untersuchen wir im Labor, ob eine unfaire Bezahlung zu Stress führt. Im Experiment interagiert ein Prinzipal (Firma) jeweils mit einem Agenten (Mitarbeiter). Der Agent produziert durch eine (sinnlose) Tätigkeit (Zählen von Nullen auf Blättern mit Nullen und Einsen) ein Einkommen, das zunächst vollumfänglich an den Prinzipal geht. Die Agenten arbeiten maximal für 25 Minuten, während die Prinzipale nicht arbeiten. Dieser Sachverhalt ist beiden Parteien bekannt. Nachdem der Agent seine Arbeit abgeschlossen und ein gemeinsames Einkommen erwirtschaftet hat, entscheidet der Prinzipal, wie er den Betrag zwischen sich und dem Agenten aufteilt. Jede Aufteilung ist möglich. Bevor der Agent die tatsächliche Aufteilung kennt, wird er gefragt, was für ihn eine angemessene Entlohnung wäre. Als biologischer Marker für Stress wird während des gesamten Experiments die Herzfrequenzvariabilität (HFV) der Agenten gemessen, wobei die HFV gering

8. Siehe hierzu, und zum Zusammenhang zwischen leistungsabhängiger Bezahlung und Risikobereitschaft, auch Grund und Sliwka (2006).

Tabelle 1 Determinanten leistungsabhängiger Bezahlung

Abhängige Variable	1 falls leistungsabhängige Bezahlung		
	(1)	(2)	(3)
Risikobereitschaft	0.027*** [0.006]	0.024*** [0.006]	0.023*** [0.007]
Reziprozität	0.009*** -0.037*** [0.014]	0.008*** -0.029** [0.014]	0.008*** -0.032** [0.015]
Vertrauen	-0.012*** 0.027* [0.014]	-0.010** 0,002 [0.014]	-0.011** -0,004 [0.016]
1 falls weiblich	0.009* -0.227*** [0.028]	0,001 -0.229*** [0.028]	-0,001 -0.065* [0.036]
Schuljahre	-0.077*** [0.028]	-0.077*** 0.041*** [0.005]	-0.021* 0.040*** [0.006]
Weitere Kontrollvariablen	keine	keine	Alter (-), Arbeitserfahrung (.), Teilzeit (-), Firmengröße (+), Arbeitsmarktregion (.)
Konstante	-0.594*** [0.037]	-1.101*** [0.073]	-1.818*** [0.133]
Pseudo R ²	-5818,57	-5682,33	-4709,2
Anzahl Beobachtungen	9813	9647	8865

Anmerkung: Probitschätzung; Standardfehler in eckigen Klammern; marginale Effekte in runden Klammern; Signifikanzniveau: *10%; **5%; ***1%

Quelle: Dohmen und Falk (2006)

ist, wenn der Stresslevel hoch ist. Epidemiologische Untersuchungen zeigen, dass HFV ein Frühindikator für Störungen des kardiovaskulären Systems mit erhöhtem Risiko für Herz-Kreislauferkrankung ist (Dekker et al., 2000; Liao et al., 1997).

Die Resultate belegen einen systematischen Einfluss einer als unfair wahrgenommenen Entlohnung und Stress. Im Durchschnitt halten die Agenten eine Aufteilung von 63% des Gesamteinkommens für sich für angemessen, erhalten aber im Schnitt nur 43%. Individuell ist die Diskrepanz zwischen angemessener oder fairer Entlohnung und tatsächlicher Entlohnung sehr unterschiedlich. Diese individuell verschiedene Variation ermöglicht es, die Hypothese zu testen, wonach unfaire Bezahlung Stress erzeugt. Abbildung 8 zeigt diesen Zusammenhang. Die Herzfrequenzvariabilität (HFV) wird abgetragen gegen die tatsächlich vom Prinzipalen vorgenommene Aufteilung. Je geringer der Anteil für den Agenten, desto geringer die HFV. Derselbe Zusammenhang ergibt

Kontrolliert und repräsentativ: Beispiele zur Komplementarität von Labor- und Felddaten

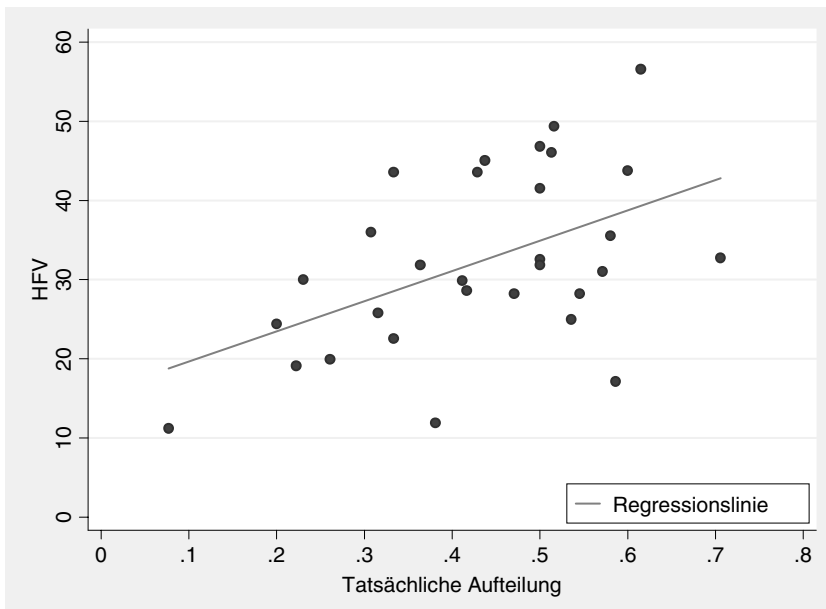


Abbildung 8 Herzfrequenzvariabilität und Entlohnung

Quelle: Falk et al. (2008)

sich auch, wenn statt der tatsächlichen Aufteilung die Diskrepanz zwischen angemessener und tatsächlicher Bezahlung untersucht wird. Die dargelegte Kausalität ist für die Interpretation von Felddaten relevant. Natürlich kann man nicht ausschließen, dass im Feld auch die umgekehrte Kausalität Geltung hat. Ziel der vorliegenden experimentellen Untersuchung war es aber, die Existenz einer bestimmten Kausalität zu untersuchen.

5. Labor- und Felddaten: Eine Liebesaffäre?

Häufig werden Laborexperimente kritisiert, sie seien nicht „realistisch“, „verallgemeinerbar“ oder „repräsentativ“. Gleichzeitig beklagen wir, wenn Resultate, die auf Felddaten basieren, kausal nicht eindeutig interpretierbar sind oder Befragungsinstrumente verwendet werden, deren Verhaltensrelevanz zumindest fraglich ist. Was liegt also näher, als Labor- und Felddaten wenn möglich zu verbinden, um die gegenseitigen Stärken zu nutzen?

An ausgewählten Beispielen haben wir versucht darzulegen, dass sich diese Zusammenarbeit lohnen kann: So ist es möglich und für viele Fragen auch sinnvoll, Experimente nicht ausschließlich mit Studenten sondern mit repräsentativen Stichproben durchzuführen oder spezifischen anderen

Personengruppen wie z.B. Kunden einer Firma (Johnson et al., 2008) oder Manager (Cooper et al., 1999). Es gibt zweitens die Möglichkeit, Fragebogen-Items mit Hilfe von Experimenten auf ihre Verhaltensrelevanz hin zu überprüfen. Hat man gezeigt, dass Antworten auf eine bestimmte Frage in der Lage sind, Verhalten unter kontrollierten Bedingungen zu prognostizieren, wächst das Vertrauen in die verwendeten Maße. Gleichzeitig können die validierten Instrumente relativ kostengünstig für große Fallzahlen administriert werden. Schließlich ergibt sich drittens durch die Kombination von Labor- und Felddaten die Möglichkeit, die Verallgemeinerbarkeit von Laborergebnissen im Feld zu überprüfen sowie gleichzeitig kausale Interpretationen von Felddaten zu erleichtern.

Es wäre meines Erachtens sinnvoll, bei der Planung von Forschungsvorhaben von Anfang an die Möglichkeiten der Verknüpfung von Labor- und Felddaten zu prüfen. Wenn dies zu einer verstärkten Kooperation von Forschern mit unterschiedlichem methodischem Hintergrund und einem verbesserten wechselseitigen Verständnis führen sollte: umso besser!

Literaturverzeichnis

- Andersen, S., G.W. Harrison, M.I. Lau und E. Rutström (2008), Eliciting Risk and Time Preferences, *Econometrica* 76, 583–618.
- Bonin, H., U. Sunde, A. Falk und T. Dohmen (2007), Cross-Sectional Earnings Risk and Occupational Sorting: The Role of Risk Attitudes, *Labour Economics* 14, 926–937.
- Borghans, L., A.L. Duckworth, J. Heckman und B. ter Weel (2008), The Economics and Psychology of Personality Traits, *Journal of Human Resources* 43, 972–1059.
- Chiappori, P.-A. und P.J. Reny (2006), Matching to Share Risk, Working Paper, Columbia University.
- Cooper, D.J., J.H. Kagel, W. Lo und Q.L. Gu, (1999), Gaming against Managers in Incentive Systems: Experimental Results with Chinese Students and Chinese Managers, *American Economic Review* 89, 781–804.
- Dekker, J.M., R.S. Crow, A.R. Falsom, P.J. Hannan, D. Liao, C.A. Swenne und E.G. Schouten (2000), Low Heart Rate Variability in a 2-Minute Rhythm Strip Predicts Risk of Coronary Heart Disease and Mortality from Several Causes: The ARIC Study, *Circulation* 102, 1239–1244.
- Dohmen, T. und A. Falk (2006), Performance Pay and Multi-dimensional Sorting: Productivity, Preferences and Gender, IZA Discussion Paper 2001.
- Dohmen, T., A. Falk, D. Huffman und U. Sunde (2006), The Intergenerational Transmission of Risk and Trust Attitudes, IZA Discussion Paper 2380.
- Dohmen, T., A. Falk, D. Huffman und U. Sunde (2007), Are Risk Aversion and Impatience Related to Cognitive Ability?, IZA Discussion Paper 2735.
- Dohmen, T., A. Falk, D. Huffman und U. Sunde (2008), Representative Trust and Reciprocity: Prevalence and Determinants, *Economic Inquiry* 46, 84–90.
- Dohmen, T., A. Falk, D. Huffman und U. Sunde (2009), Homo Reciprocans: Survey Evidence on Behavioural Outcomes, *Economic Journal* 119, 592–612.

Kontrolliert und repräsentativ: Beispiele zur Komplementarität von Labor- und Felddaten

- Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp und G.G. Wagner (2005), Individual Risk Attitudes: New Evidence from a Large, Representative, Experimentally-Validated Survey, IZA Discussion Paper 1730.
- Falk, A. und J. Heckman (2009), The laboratory is a major source of knowledge in economics, Working Paper, Universität Bonn.
- Falk A., I. Menrath, P.E. Verde und J. Siegrist (2008), Unfair Pay and Cardiovascular Function, Working Paper, Universität Bonn.
- Fehr, E., U. Fischbacher, B. v. Rosenblatt, J. Schupp und G.G. Wagner (2002), A Nationwide Laboratory: Examining Trust and Trustworthiness by Integrating Behavioral Experiments into Representative Surveys, *Schmöllers Jahrbuch* 122(4), 519–54.
- Frick, J., J. Schupp und G.G. Wagner (2007), The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancement, *Journal of Applied Social Studies* 127, 139–169.
- Grund, C. und D. Sliwka (2006), Performance Pay and Risk Aversion, IZA Discussion Paper 2012
- Jaeger D. A., H. Bonin, T. Dohmen, A. Falk, D. Huffman und U. Sunde (2009) Direct Evidence on Risk Attitudes and Migration, *Review of Economics and Statistics*, im Erscheinen
- Johnson, E.J., S. Gächter und A. Herrmann (2006), Exploring the Nature of Loss Aversion, CeDEX Discussion Paper 2006–02.
- Lang, F.R. (2005), Erfassung des kognitiven Leistungspotenzials und der “Big Five” mit Computer-Assisted-Personal-Interviewing (CAPI): zur Reliabilität und Validität zweier ultrakurzer Tests und des BFI-S, Research Note No. 9, DIW Berlin.
- Liao, D., J. Cai, W.D. Rosamond, R.W. Barnes, R.G. Hutchinson, E.A. Whitsel, P. Rautaharju und G. Heiss (1997), Cardiac Autonomic Function and Incident Coronary Heart Disease: A Population-Based Case-Cohort Study: The ARIC Study: Atherosclerosis Risk in Communities Study, *American Journal of Epidemiology* 145, 696–706.
- Persico, N., A. Postlewaite, und D. Silverman (2004), The Effect of Adolescence. Experience on Labor Market Outcomes: The Case of Height, *Journal of Political Economy* 112, 1019–1053.
- Schupp, J. und G.G. Wagner (2002), Maintenance of and Innovation in Long-Term Panel Studies The Case of the German Socio-Economic Panel (GSOEP), *Allgemeines Statistisches Archiv* 86, 163–175.
- Sichelstiel, G. und F. Söllner (2004), Gleich und gleich gesellt sich gern: Ökonomische Ansätze zur Partnerwahl, *Perspektiven der Wirtschaftspolitik* 5, 249–270.
- Siegrist, J. (1996), Adverse Health Effects of High Effort/ Low Reward Conditions, *Journal of Occupational Health Psychology* 1, 27–41.
- Wagner, G., Goebel, J., Krause, P., Pischner, R. und Sieber, I. (2008), Das Sozio-oekonomische Panel (SOEP): Multidisziplinäres Haushaltspanel und Kohortenstudie für Deutschland, *ASTa Wirtschafts- und Sozialstatistisches Archiv* 2, 301–328.

Abstract: *In this article we discuss the complementarity of laboratory and field data. Experiments offer highly controlled environments that allow precise testing and causal inferences. Survey and field data on the other hand provide information on large and representative samples of people interacting in their natural environment.*

We discuss several concrete examples how to combine lab and field data and how to exploit potential complementarities. One example describes an experiment, which is run with a representative sample to guarantee control and representativeness. The second example is based on the idea to experimentally validate survey instruments to ensure behavioral validity of instruments that can be used in existing panel data sets. The third example describes the possibility to use the lab to identify causal effects, which are tested in large data sets. Topics discussed in this article comprise the relation of cognitive skills (IQ) and risk and time preferences, determinants, prevalence and economic consequences of risk attitudes, selection into incentive schemes and the impact of unfair pay on stress.